Phase Diagrams of Quasispecies Theory with Recombination and Horizontal Gene Transfer

J.-M. Park[1,2] and M. W. Deem[1]

[1] *Department of Physics & Astronomy, Rice University, Houston, Texas 77005–1892, USA*
[2] *Department of Physics, The Catholic University of Korea, Bucheon, 420–743, Korea*

We consider how transfer of genetic information between individuals influences the phase diagram and mean fitness of both the Eigen and the parallel, or Crow-Kimura, models of evolution. In the absence of genetic transfer, these physical models of evolution consider the replication and point mutation of the genomes of independent individuals in a large population. A phase transition occurs, such that below a critical mutation rate an identifiable quasispecies forms. We show how transfer of genetic information changes the phase diagram and mean fitness and introduces metastability in quasispecies theory, via an analytic field theoretic mapping.

PACS numbers: 87.10.+e, 87.15.Aa, 87.23.Kg, 02.50.-r

We consider how quasispecies evolution changes in the presence of transfer of genetic information between individuals in a population. That is, we quantify by quasispecies theory the mutational load, if any, introduced by a model of recombination and gene transfer. Exchange of genetic information between individuals is believed to be pervasive in nature and crucial to evolutionary dynamics (for reviews, see [1, 2, 3]). Experiments and theory have emphasized that recombination and gene transfer in various forms increase the rate of laboratory directed protein evolution [4, 5, 6] (for reviews see [7, 8]). Other experiments have amplified this point and have also suggested that, while significant in practice, the advantage of recombination may simply be to speed up the evolutionary process that would naturally occur by mutation alone in the limit of a long enough evolutionary time or a large enough population size [9, 10, 11].

The Eigen [12] and Crow-Kimura [13], or parallel, models of viral quasispecies evolution are among the simplest that capture the basic processes of mutation, selection, and replication that occur in natural evolution. These mathematical models exhibit phase transitions, such that for mutation rates below critical values, an identifiable quasispecies forms. The Eigen and parallel quasispecies models are archetypes of biological evolution, and they have become a popular entry point to evolutionary biology for physicists [14, 15, 16, 17, 18, 19, 20, 21]. Quantification of the mutational load of transfer of genetic information has been done by numerical solutions of the single-mutation-per-replication Eigen model for the special case of a linear replication rate function [22]. It was found that for intermediate population sizes and finite times, genetic transfer dramatically speeds up the rate of evolution. Phase diagrams were not determined, due to the focus on finite times and population sizes. We here derive by analytical calculation the mutational load and evolutionary advantage induced by transfer of genetic information for arbitrary replication rate functions in both the parallel and continuous-time Eigen models of quasispecies theory. That is, we find the infinite-time, infinite-genome-length, and infinite-population-size phase diagrams and mean fitness values of these models of quasispecies evolution in the general case. As an example, for the sharp-peak replication rate function, transfer of genetic information has two effects: sharpening of the population in the selected phase ($u = 1$ instead of $0 < u \leq 1$) and maintaining the unselected phase as metastable to higher growth rates.

We model transfer of genetic information by replacement of part of the genome in a random individual with sequence taken at the same genomic location from a random parent. One common way for this process to occur in viruses or bacteria is by recombination. We assume that each of these exchanges of genetic information changes only one base of the sequence, such as might occur by homologous recombination in a stable population with relatively low diversity. We consider an infinite population of individuals with fixed genome size N, and each site of the genome may be in one of two states (e.g. purine or pyrimidine). A given sequence i reproduces at rate r_i, and point mutations occur at rate μ per site to change sequence j to i.

Transfer of genetic information occurs in the process in which a base at position k from any sequence j randomly replaces the base at position k in sequence i' with frequency ν. Quasispecies theory with transfer of genetic information is described by the equation

$$\frac{dq_i}{dt} = r_i q_i + \sum_{j=1}^{2^N} \mu_{ij} q_j + \nu \frac{\sum_{k=1}^{N} \sum_{i'}' \sum_j' q_{i'} q_j}{\sum_{m=1}^{2^N} q_m} - \nu N q_i$$

$$= r_i q_i + \sum_{j=1}^{2^N} \mu_{ij} q_j + \nu \sum_{k=1}^{N} (q_i + q_{\sigma_1(k)i}) \frac{\sum_j' q_j}{\sum_{m=1}^{2^N} q_m} - \nu N q_i \ , \quad (1)$$

where the primes indicate that sequence j equals sequence i at position k, and sequence i' equals sequence i except possibly at position k. The notation $i' = \sigma_1(k)i$ indicates the sequence i' that results from changing base k in sequence i. We have defined q_i to be proportional to the probability of sequence i in the population. We note that the recombination term conserves particle number: taking the sum of Eq. (1) over i causes the mutation and recombination terms to cancel. We define the average spin at position k at time t as $u(k)$, $u(k) = \sum_j (\delta_{s_k^j,+1} - \delta_{s_k^j,-1}) q_j / \sum_m q_m$, where s_k^j represents the base at position k of sequence j. The recombination process is, thus, described by

$$\frac{dq_i}{dt} = r_i q_i + \mu \sum_{k=1}^{N} \left(q_{\sigma_1(k)i} - q_i \right) + \nu \sum_{k=1}^{N} (q_i + q_{\sigma_1(k)i}) \times \left(\frac{1+u(k)}{2} \delta_{s_k^i,+1} + \frac{1-u(k)}{2} \delta_{s_k^i,-1} \right) - \nu N q_i \ . \quad (2)$$

We analyze this equation at long times, when u becomes independent of both time and position. In this limit, the non-linear Eq. (2) becomes a linear equation, with a self-consistency condition for u. We find the long-time solution by a mapping to a two-component field theory, using the procedure of [17]. We define the space $|\psi(t)\rangle = \sum_i^{2^N} q_i(t) |s^i\rangle$. The evolution equation is then cast as $\frac{d}{dt}|\psi\rangle = -\hat{H}|\psi\rangle$. The space is represented in terms of creation and annihilation operators. We seek the operator form of Eq. (1). Each spin state, $|s_n^i\rangle = |+1\rangle$, $|-1\rangle$, is created by a different operator.

We introduce two pairs of creation and annihilation operators for the space, with the constraint that at each position one and only one particle is present. We label the creation operators at position j as $\hat{\vec{a}}^\dagger(j) = (\hat{a}_1^\dagger(j), \hat{a}_2^\dagger(j))$. The Hamiltonian is given by

$$-\hat{H} = \mu \sum_{j=1}^{N} [\hat{M}(j) - 1] + \nu N \hat{R} + N f \left[\frac{1}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot \sigma_3 \hat{\vec{a}}(j) \right] \ , \quad (3)$$

where μ is the mutation rate, $\hat{M}(n)$ mutates s^i at position n, $Nf(u_i) = r_i$ is the replication rate, νN is the recombination rate, and $\hat{R}$ is the recombination operator. The mutation operator $M(j)$ operates on site j and are defined by $\hat{M}(j) = \hat{\vec{a}}^\dagger(j) \cdot \sigma_1 \hat{\vec{a}}(j)$. The recombination operator is given by

$$\hat{R} = \frac{1}{N} \sum_{j=1}^{N} [a_1^\dagger(j)[a_1(j) + a_2(j)] \frac{1+u}{2} + a_2^\dagger(j)[a_1(j) + a_2(j)] \frac{1-u}{2} - 1] \ . \quad (4)$$

The formal solution is $|\psi(t)\rangle = e^{-\hat{H}t}|\psi(0)\rangle$, which implies that the joint probability distribution at time t is given by $q_i(t) = \langle s_i | e^{-\hat{H}t} \sum_l q_l(0) | s^l \rangle$. We introduce the coherent state representation $|\vec{z}\rangle = e^{\hat{\vec{a}}^\dagger \cdot \vec{z} - \vec{z}^* \cdot \hat{\vec{a}}} |0,0\rangle$, where $\vec{z} = (z_1, z_2)$ is a two-vector. The coherent states satisfy a completeness relation $I = \int [\mathcal{D}\vec{z}^* \mathcal{D}\vec{z}] |\vec{z}\rangle\langle\vec{z}|$. The overlap of the coherent states is given by $\langle \vec{z}' | \vec{z} \rangle = e^{-\frac{1}{2}[\vec{z}'^* \cdot (\vec{z}' - \vec{z}) - (\vec{z}'^* - \vec{z}^*) \cdot \vec{z}]}$. Using [17], we use the Trotter factorization to find the evolution operator is

$$e^{-\hat{H}t} = \lim_{M \to \infty} \int [\prod_{k=0}^{M} \mathcal{D}\vec{z}_k^* \mathcal{D}\vec{z}_k] |\vec{z}_M\rangle \left[\prod_{k=1}^{M} \langle \vec{z}_k | e^{-\varepsilon \hat{H}} | \vec{z}_{k-1} \rangle \right] \langle \vec{z}_0 | \ , \quad (5)$$

where $\varepsilon = t/M$.

The probability to go from an initial state $q_{(i)=\gamma}$ to a final state $q_{(i')=\gamma'}$, where $1 \le \gamma_j \le 2$ indicates the composition of the pair of bases at position j, is

$$P = \lim_{M \to \infty} \int \mathcal{D}\bar{\xi}^* \mathcal{D}\xi e^{\varepsilon N \sum_{k=1}^{M} [f(\xi_k) - \bar{\xi}_k \xi_k]} \prod_{j=1}^{N} Q_{\gamma_j' \gamma_j}(j) \ , \quad (6)$$

where $Q(j) = \prod_{k=1}^{M} [I + \varepsilon B_k(j)]$, with $B_k(j) = \mu(\sigma_1 - I) + \nu(D - I) + \bar{\xi}_c \sigma_3$, and $D = \begin{pmatrix} \frac{1+u}{2} & \frac{1+u}{2} \\ \frac{1-u}{2} & \frac{1-u}{2} \end{pmatrix}$.

We are interested in the probability distribution at long times, which for a given u grows as $e^{f_m t}$ by the Perron-Frobenius theorem, where f_m is the largest eigenvalue of $-\hat{H}$, and f_m is equal to the mean replication rate at long times [17], when u is self-consistently determined. We evaluate the contribution to this eigenvalue from $Q(j)$ by considering the expression Tr $Q(j)$. We find

$$\ln \text{Tr } Q(j) \sim t \left(\sqrt{(\mu + \nu/2)^2 + \nu u \bar{\xi}_c + \bar{\xi}_c^2} - \mu - \nu/2 \right) \ . \quad (7)$$

We note that in the limit of infinite ν, $\ln \text{Tr } Q(j) \sim t u \bar{\xi}_c$. The mean replication rate is given by $f_m = \max_{\xi_c, \bar{\xi}_c} \{ f(\xi_c) - \bar{\xi}_c \xi_c + [\ln \text{Tr } Q(j)]/t \}$. Maximizing over $\bar{\xi}_c$, we find

$$f_m = \max_{\xi_c} \left\{ f(\xi_c) + \left[(\mu + \nu/2)^2 - (\nu u/2)^2 \right]^{1/2} \sqrt{1 - \xi_c^2} + \nu u \xi_c / 2 - \mu - \nu/2 \right\} \ . \quad (8)$$

The observable surface magnetization, u, is given by the implicit self-consistency condition $f(u) = f_m$. Thus, the

two variables ξ_c and u need to be determined when solving Eq. (8). This procedure provides the exact solution to the parallel model of recombination for a general replication rate function.

To illustrate how recombination affects the error-threshold phase transition, we calculate the error threshold for three different replication rate functions. We first consider in detail $f(1) = A$ and $f = 0$ otherwise. Eq. (8) is maximized at $\xi_c = 1$ or $\xi_c = 0$. The error threshold is given for $u = 0$ by $A > \mu + \nu/2$. The self-consistency condition $f_{\rm m} = f(u)$ can only be satisfied by $u = 1 - O(1/N)$. Thus, due to the non-linearity, $u = 1$ in the selected phase, in contrast to the case without recombination, for which $u = 1 - \mu/A$ [17]. In the selected phase $f_{\rm m} = A - \mu$. Thus, the true error threshold is $A > \mu$, with $A > \mu + \nu/2$ the limit of metastability for initial conditions with $u \approx 0$. If we are in the selected phase and reduce the replication rate of the sharp peak, we transform to the unselected phase at the solid line of Figure 1. If, however, we are in the unselected phase and increase the replication rate of the sharp peak, we may not transform to the selected phase until reaching the short-dashed line of Figure 1. We next consider in detail the case of the quadratic replication rate $f(u) = ku^2/2$. By setting Eq. (8) equal to $f(u)$ for small u, we find that the error threshold is given by $k > (\mu + \nu)/[1 + \nu/(2\mu)]$. At small ν, recombination has again shifted the transition by $+\nu/2$. As an example of general parameter values, for $\nu = 1$, $\mu = 1$, and $k = 2$, we find $u = 0.4671$ and $f_{\rm m} = 0.2182$. Solving Eq. (1) numerically for $N = 10^3$, $\nu = 1$, $\mu = 1$, and $k = 2$, we find $u = 0.4662$ and $f_{\rm m} = \langle f(u_l) \rangle = 0.2183$. Note recombination introduces a genetic load for the quadratic replication rate, since in the absence of recombination $u = 1 - \mu/k = 1/2$ for these parameters. Note also that metastability does not occur for the quadratic replication rate. The error-threshold phase diagram for these two cases is shown in Figure 1. We finally consider in detail the linear fitness $f(\xi) = k_0 + k\xi$. We find that for all values of k_0, $k > 0$, $\mu \geq 0$, and $\nu \geq 0$, the optimal value of ξ_c is positive. Thus, the selected phase always occurs

We now turn to consider recombination in the Eigen model. In the Eigen model, when a virus reproduces, the virus copies its genome, making mutations at a rate of $1 - y$ per base per replication. The un-normalized probability distribution in genome space satisfies

$$\frac{dq_i}{dt} = \sum_{j,k=1}^{2^N} \left[B_{ij} C_{jk} r_k - \delta_{ij} \delta_{ik} D_i \right] q_k \ . \tag{9}$$

The degradation rate is defined analogously to the replication rate by $D_i = Nd(u_i)$. Here the transition rates are given by $B_{ij} = y^{N-d(i,j)}(1-y)^{d(i,j)}$. We define the parameter $\mu = N(1-y)/y$ to characterize the per genome replication rate, where we take $\mu = O(1)$, consistent with observed mutation rates in many viruses and bacteria

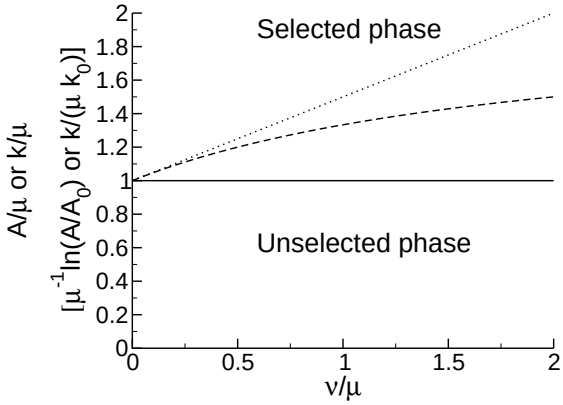

FIG. 1: The selected phase, in which a finite fraction of the population has a non-zero magnetization, is shown for the parallel model with recombination. The rate of genetic exchange is ν, and the rate of mutation is μ. The phase diagram is shown for the replication rate $f(u) = A\delta_{u,1}$ (solid line, with short-dashed line the metastable limit) and the replication rate $f(u) = ku^2/2$ (dashed line). Also shown is the phase diagram for the Eigen model for $f(u) = (A - A_0)\delta_{u,1} + A_0$ (solid line, with short-dashed line the metastable limit) and $f(u) = ku^2/2 + k_0$ (long-dashed line) [ordinate coordinates in brackets].

[23]. We also have $C_{jk} \sim \Big[-\nu + (\nu/N) \sum_{l=1}^{N} (\delta_{jk} + \delta_{\sigma_1(l)j,k}) \left(\frac{1+u(l)}{2} \delta_{s_l^j,+1} + \frac{1-u(l)}{2} \delta_{s_l^j,-1} \right) \Big]$. The rate of genetic exchange per site is ν/N. This equation is generated to $O(N^0)$ by the Hamiltonian

$$\begin{aligned} -\hat{H} \ = \ & N e^{-\mu + \frac{\mu}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot \sigma_1 \hat{\vec{a}}(j)} \times e^{-\nu + \frac{\nu}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot D \hat{\vec{a}}(j)} \\ & \times f \left[\frac{1}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot \sigma_3 \hat{\vec{a}}(j) \right] \\ & - N d \left[\frac{1}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot \sigma_3 \hat{\vec{a}}(j) \right] \ . \end{aligned} \tag{10}$$

We define $\chi_k = \frac{1}{N} \sum_{j=1}^{N} \hat{\vec{a}}^\dagger(j) \cdot D \hat{\vec{a}}(j)$ and $\eta_k = \frac{1}{N} \sum_{j=1}^{N} \vec{z}_k^*(j) \sigma_1 \vec{z}_{k-1}(j)$. We integrate out the $\mathbf{z}$ field, to find $B_k(j) = \bar{\eta}_k \sigma_1 + \bar{\xi}_k \sigma_3 + \bar{\chi}_k D$. The action is, therefore, given to $O(N^0)$ by

$$\begin{aligned} -S \ = \ & \varepsilon N \sum_{k=1}^{M} \Big[e^{-\mu + \mu\eta_k - \nu + \nu\chi_k} f(\xi_k) - d(\xi_k) \\ & - \bar{\xi}_k \xi_k - \bar{\eta}_k \eta_k - \bar{\chi}_k \chi_k \Big] + N \ln \mathrm{Tr} \ Q(j) \ , \end{aligned} \tag{11}$$

where $\ln \mathrm{Tr} \ Q(j) \sim t(\sqrt{(\bar{\eta}_c + \bar{\chi}_c/2)^2 + u\bar{\chi}\bar{\xi}_c + \xi_c^2}) + \bar{\chi}_c/2$. Note that the probability of any of the N bases undergoing both mutation and recombination is $O(\nu\mu/N)$. We have $f_{\rm m} = \max_{\xi_c, \bar{\xi}_c, \eta_c, \bar{\eta}_c, \chi_c, \bar{\chi}_c} \{ e^{-\mu + \mu\eta_c - \nu + \nu\chi_c} f(\xi_c) -$

$d(\xi_c) - \bar{\xi}_c\xi_c - \bar{\eta}_c\eta_c - \bar{\chi}_c\chi_c + [\ln \mathrm{Tr}\ Q(j)]/t\}$. Maximizing over $\bar{\xi}_c$, $\bar{\eta}_c$, and $\bar{\chi}_c$, we find that $\bar{\eta}_c\eta_c + \bar{\chi}_c\chi_c + \bar{\xi}_c\xi_c = (\ln q)/t$. We use the additional relation $\mu\bar{\chi}_c = \nu\bar{\eta}_c$ to find $\eta_c(\xi_c) = \{(1-\xi_c^2)/[1-\nu^2u^2/(2\mu+\nu)^2]\}^{1/2}$ and $\chi_c(\xi_c) = \left[1 + \eta_c + u\xi_c - u\left(\eta_c^2 + \xi_c^2 - 1\right)^{1/2}\right]/2$. The mean replication rate is given by the expression

$$f_{\mathrm{m}} \ = \ \max_{\xi_c} \left\{ e^{-\mu+\mu\eta_c-\nu+\nu\chi_c} f(\xi_c) - d(\xi_c) \right\} \ . \quad (12)$$

The observable u is given implicitly by $f_{\mathrm{m}} = f(u) - d(u)$.

To illustrate how recombination during replication affects the error-threshold phase transition, we calculate the error threshold for three different replication rate functions. We first consider in detail $f(1) = A$ and $f = A_0$ otherwise. Eq. (12) is maximized at $\xi_c = 1$ or $\xi_c = 0$. The error threshold is given for $u = 0$ by the equation $Ae^{-\mu-\nu/2} > A_0$. Due to the non-linearity, $u = 1$ in the selected phase, in contrast to the case without recombination where $u = (Ae^{-\mu} - A_0)/(A - A_0)$ [17]. The mean replication rate is given by $f_{\mathrm{m}} = Ae^{-\mu}$. Thus, the true error threshold is $Ae^{-\mu} > A_0$, with $Ae^{-\mu-\nu/2} > A_0$ the limit of metastability for initial conditions with $u \approx 0$. The limits of the bistable region, which is reminiscent of the bistable region found numerically in a related Eigen model with a type of recombination [24], are demarked by the solid and short-dashed lines in in Figure 1. For our second example, we consider the quadratic fitness $f(\xi) = k_0 + k\xi^2/2$. By setting Eq. (12) equal to $f(u)$ for small u, we find that the error threshold is given by $k > k_0(\mu+\nu)/[1+\nu/(2\mu)]$. At small ν, recombination has again shifted by transition by $+\nu/2$. The error-threshold phase diagram for the sharp peak and quadratic replication rate cases is shown in Figure 1. For our third example, we consider in detail the linear fitness $f(\xi) = k_0 + k\xi$. We find that for all values of $k > 0$, $\mu \geq 0$, and $\nu \geq 0$, the optimal value of ξ_c is positive, and the selected phase always occurs.

How is it that recombination changes the phase diagram or mean fitness? After all, this process simply replaces an allele with another allele randomly chosen from the distribution of alleles at that site. This process, however, reduces the correlations between the composition of alleles at different sites in the sequence. These correlations are non-zero [17], and so their reduction changes the dynamics and the steady-state distribution. Recombination sharpens the phase transition for $f(\xi) = f(0) + \delta_{\xi,1}\Delta A$, turning it into step function for u. More generally, recombination propagates favorable mutations throughout the population, thereby typically increasing the rate of evolution.

We may alternatively interpret each site in our model as a coarse-grained representation of an allele or gene, each with only two states that may be changed either by mutation or gene transfer. We may also consider that the genome has been ordered so that all the mobile genetic elements are collected, say, at the end. If the genome can be considered approximately constant in length and if the replication rate can be approximately expressed by the quasispecies assumption $r_i = Nf(u_i)$, then the model we have discussed is a representation of horizontal gene transfer in a population of evolving bacterial species, because homology is not a prerequisite for horizontal gene transfer in nature or our model. The diversity in the population represents the species diversity in a bacterial order, family, or genus. As with natural gene transfer by mobile elements, our model does not assume homology is required. Our results make a couple generic predictions: 1) for a sharp peak fitness, such as might be induced by a novel antibiotic, a population with horizontal gene transfer tends to be more uniformly resistant ($u \approx 1$) than is a population without ($0 < u < 1$), 2) while the contributions of mutation and horizontal gene transfer to the mean fitness are not identical, in Eq. (8) or (12), they are similar for smooth replication rates. Horizontal gene transfer tends to incorporate alleles with new function, whereas mutation tends to adapt existing alleles for improved function. The observed rates of horizontal gene transfer and mutation might be expected to be the same order of magnitude, therefore, to balance the resources expended on the complementary tasks of large-scale evolution and local adaptation. As an example, we consider the evolution of *E. coli* from *Salmonella*. The rate of evolution due to horizontal gene transfer is estimated to be 16,000 bases/million years, while that due to point mutation, is 22,000 bases/million years [25]. These are observed rates, and so selection plays a role. The underlying rate of horizontal gene transfer in *E. coli* has been estimated to be about 10^{-6} genes per cell per replication [26], which corresponds to a change of roughly 10^{-3} bases per sequence per replication, given the average *E. coli* gene length of 10^3 bases. Taking the typical underlying *E. coli* mutation rate of 5×10^{-10} per base per replication [23] and noting that the *E. coli* genome length is $\approx 5 \times 10^6$ bases, we find that point mutation modifies approximately 2.5×10^{-3} bases per sequence per replication. Interestingly, this same equality of underlying horizontal gene transfer and point mutation rates per base per replication is also observed in quantitative models of laboratory directed protein evolution optimized for evolutionary rate [6].

[1] J. G. Lawrence, Trends Microbiol. **5**, 355 (1997).
[2] J. A. Shapiro, J. Biol. Phys. **28**, 745 (2002).
[3] S. P. Otto and T. Lenormand, Nature Rev. Gen. **3**, 252 (2002).
[4] W. P. C. Stemmer, Nature **370**, 389 (1994).
[5] A. Crameri, S. A. Raillard, E. Bermudez, and W. P. C. Stemmer, Nature **391**, 288 (1998).
[6] L. D. Bogarad and M. W. Deem, Proc. Nat. Acad. Sci. USA **96**, 2591 (1999).

[7] P. A. Patten, R. J. Howard, and W. P. C. Stemmer, Curr. Opin. Biotech. **8**, 724 (1997).

[8] S. Lutz and S. J. Benkovic, Curr. Opin. Biotech. **11**, 319 (2000).

[9] J. Maynard and G. Georgiou, Ann. Rev. Biomed. Eng. **2**, 399 (2000).

[10] P. D. Holler, P. O. Holman, E. V. Shusta, S. O'Herrin, K. D. Wittrup, and D. M. Kranz, Proc. Nat. Acad. Sci. USA **97**, 5387 (2000).

[11] J. Hanes and A. Plückthun, Proc. Nat. Acad. Sci. USA **94**, 4937 (1997).

[12] M. Eigen, Naturwissenschaften **58**, 465 (1971).

[13] J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper and Row, New York, 1970).

[14] K. Jain and J. Krug, in *Structural approaches to sequence evolution: Molecules, networks and populations*, edited by U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo (Springer Verlag, Berlin, 2006), q-bio.PE/0508008.

[15] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997), **79**, 1782.

[16] D. B. Saakian and C.-K. Hu, Phys. Rev. E **69**, 021913 (2004).

[17] J.-M. Park and M. W. Deem, J. Stat. Phys. (2006), dOI:10.1007/s10955-006-9190-z, q-bio.PE/0607012.

[18] D. B. Saakian, E. Munoz, C.-K. Hu, and M. W. Deem, Phys. Rev. E **73**, 041913 (2006).

[19] I. Leuthäusser, J. Chem. Phys **84**, 1884 (1986).

[20] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).

[21] L. Peliti, Europhys. Lett. **57**, 745 (2002).

[22] E. Cohen, D. A. Kessler, and H. Levine, Phys. Rev. Lett. **94**, 098102 (2005).

[23] J. W. Drake, Proc. Natl. Acad. Sci. USA **88**, 7160 (1991).

[24] M. C. Boerlijst, S. Bonhoeffer, and M. A. Nowak, Proc. Roy. Soc. Lond. B **263**, 1577 (1996).

[25] J. G. Lawrence and H. Ochman, Proc. Natl. Acad. Sci. USA **95**, 9413 (1998).

[26] C. G. Kurland, B. Canback, and O. G. Berg, Proc. Natl. Acad. Sci. USA **100**, 9658 (2003).